Returns to higher education subjects and tiers in China

Kang, Lili; Peng, Fei; Zhu, Yu

Published in:
Studies in Higher Education

DOI:
10.1080/03075079.2019.1698538

Publication date:
2021

Document Version
Peer reviewed version

Appendix: Inverse Probability Weighted Regression Adjustment (IPWRA)

We explore the "doubly robust" IPWRA estimator (see Wooldridge 2007, Wooldridge 2010 chapter 13, and Imbens and Wooldridge 2009). Compared to the more conventional propensity score matching (PSM) method which only allows a single treatment, the IPWRA estimates the average treatment effect (ATE) of any HE type and allows for selection into a particular HE type by using multinomial logit model in the first step, which is labelled as the treatment equation. In the second step, this estimator then estimates an OLS regression of log earnings by using the reweighted data, using the inverse of the predicted probabilities from the first step as the weights. In other words, the IPWRA weights observations that are more likely to contain missing data (or in the sparse parts of the distribution) *more* heavily, to account for the missing-ness process. If the functional form is correct, then the OLS estimates are unaffected by any weighting. However, the true functional form is rarely known in practice, and any misspecification is likely to yield biased estimates.

IPWRA is *doubly* robust in the sense that only one of the two steps must be specified correctly. In other words, the estimates of the second step, the log earnings equation, are robust to misspecification in the weighting of the data conditional on the specification of the second step being correct; *and* the estimates of the second step are also robust to misspecification of the second step provided the multinomial logit weighting in the first step is correctly specified.[1] Under the conditional independence assumption, which implies there is only selection on observables, this weighting method can yield causal estimates of the average treatment effects (ATEs).

The overlap plots for men and women, respectively, in Figures A2 and A3 suggest that for graduates of all types (i.e., the treated) we observe good matches from the control group of non-HE graduates.[2]

Tables A1 and A2 in the Appendix present the full set of IPWRA results for men and women, respectively. Consistent with the literature, mother's education

[1] Belfield et al (2018) and Walker and Zhu (2018) are two recent examples of the application of the IPWRA approach to estimate HE returns in the UK.

[2] We have also conducted post-estimation balancing tests which indicate that our model significantly improved the level of balance, in the sense that the weighted standardized differences are all close to zero and the variance ratios are all close to one. However, the full tables are too long to be included in the paper (but are available upon request).

qualifications seem to play the most significant role in the choice of respondents' HE types in the first step.

Table A3 tabulates the choice of HE subjects and HE tiers by mother's education, and separately for agricultural and non-agricultural *hukou* holders as defined at age 12. The patterns indicate a very strong and monotonic relationship between mother's educational qualifications and respondent's HE tiers, at least for *urban hukou* holders (notably, very few agricultural *hukou* holders had highly educated mothers). For instance, as the mother's education increases from illiteracy to university degree, the chance of an urban resident not going to HE decreases from 67% to 17% and that for going to the most selective Key Universities increases from 4% to 23%. There is also suggestive evidence that rural and urban residents might have different preferences regarding HE subjects, with the former more likely to choose STEM and the latter more inclined to choose LEM. These patterns are consistent with the literature that suggests that students with advantaged backgrounds from urban areas are over-represented in the most selective universities in China (Feng 2012).[3]

[3] IPWRA does not require exclusion restrictions as in the case of the Instrumental Variable approach. However, our findings would remain effectively unchanged if we exclude *hukou* status or mother's education from the outcome equation.

Appendix Figures

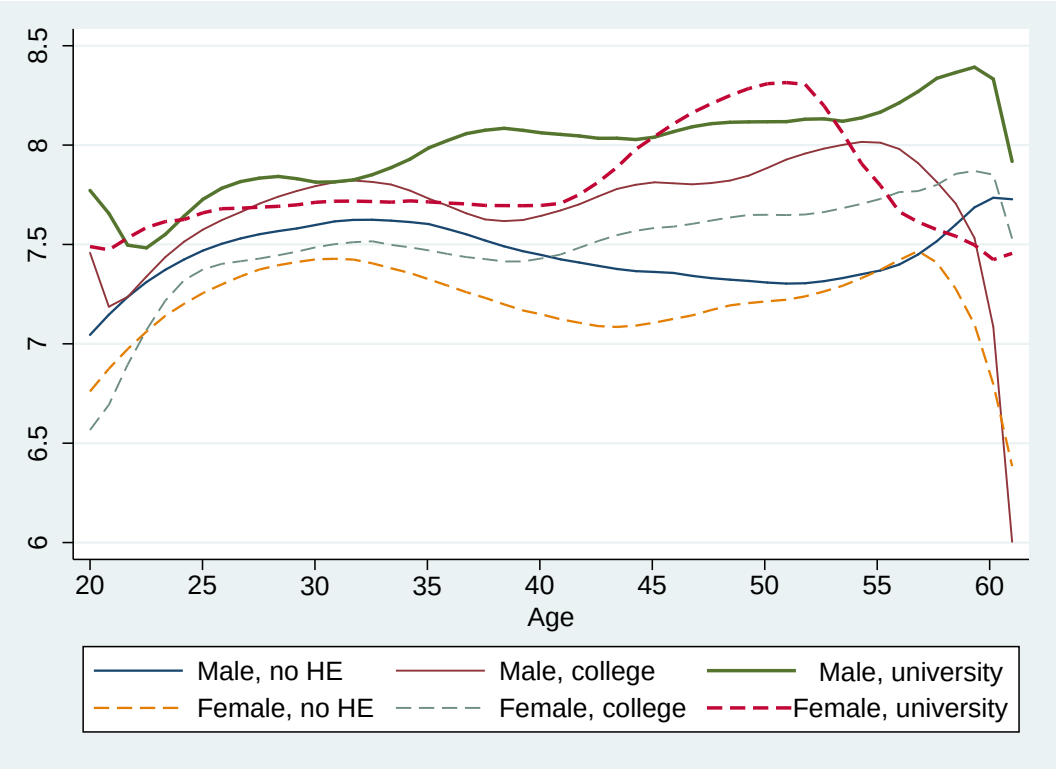

Figure A1: Smoothed age-earnings profiles by level of qualifications and gender

Note: Smoothed age-earnings profiles derived from a kernel-weighted local polynomial regression of log real monthly incomes (in January 2009 prices) on age by level of qualifications and gender, using panel sample.

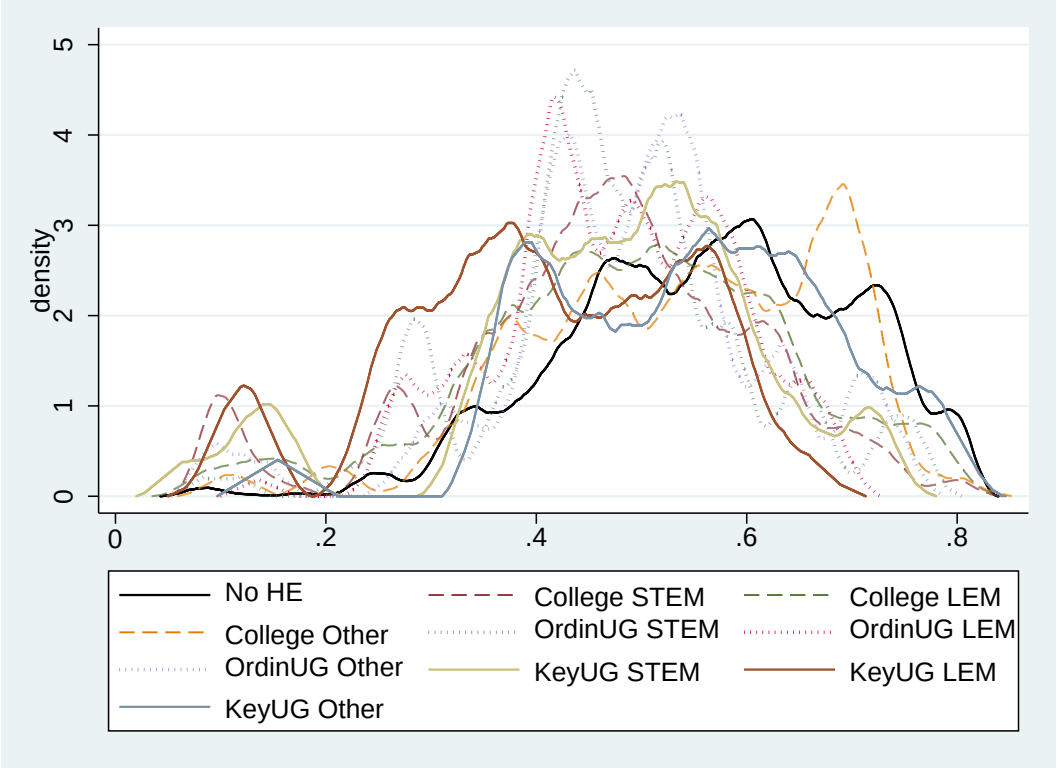

Figure A2: Overlap Plots, Men

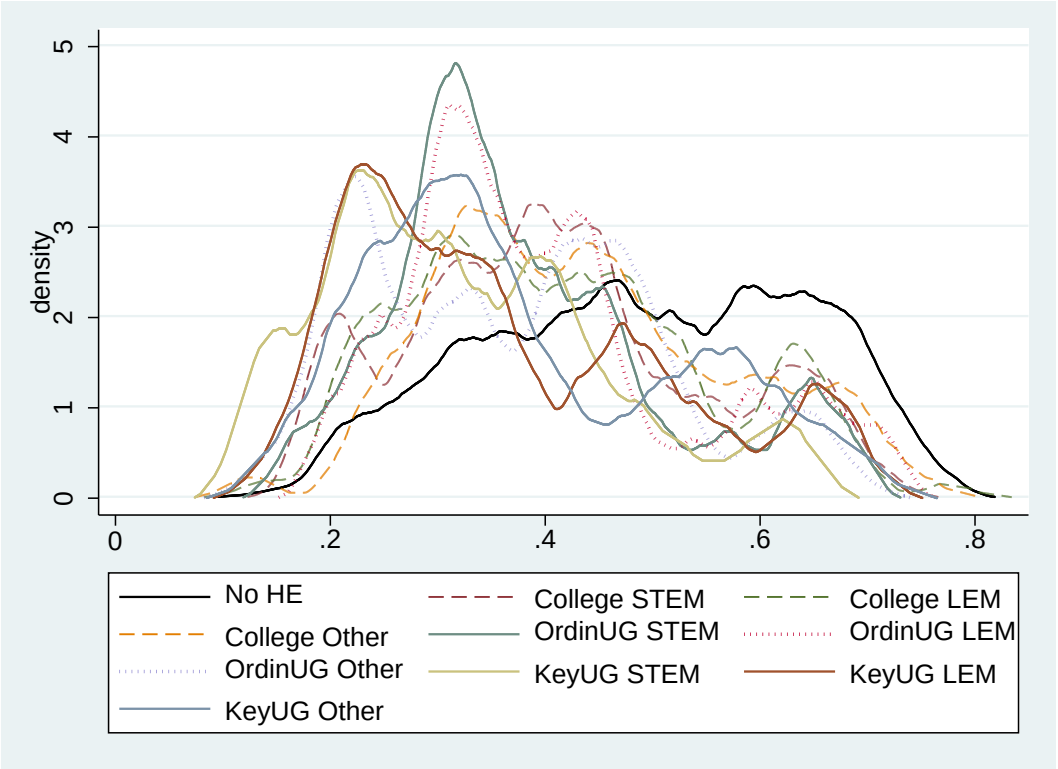

Figure A3: Overlap Plots, Women

Appendix Tables

Table A1: RE estimates by gender, with additional controls

	Men		Women	
	Mother's Year of Birth	Mother's Year of Birth + Occupation	Mother's Year of Birth	Mother's Year of Birth + Occupation
	(1)	(2)	(3)	(4)
College STEM	0.234***	0.180***	0.222***	0.175***
	(0.048)	(0.048)	(0.061)	(0.061)
College LEM	0.297***	0.258***	0.228***	0.204***
	(0.044)	(0.045)	(0.046)	(0.046)
College Other	0.275***	0.177**	0.290***	0.226***
	(0.070)	(0.074)	(0.072)	(0.074)
OrdinUG STEM	0.331***	0.239***	0.513***	0.439***
	(0.087)	(0.089)	(0.086)	(0.090)
OrdinUG LEM	0.386***	0.324***	0.406***	0.385***
	(0.088)	(0.088)	(0.103)	(0.103)
OrdinUG Other	0.408***	0.309***	0.437***	0.343***
	(0.081)	(0.086)	(0.082)	(0.087)
KeyUG STEM	0.644***	0.551***	0.610***	0.518***
	(0.086)	(0.087)	(0.148)	(0.148)
KeyUG LEM	0.619***	0.536***	0.578***	0.530***
	(0.114)	(0.112)	(0.094)	(0.098)
KeyUG Other	0.488***	0.399***	0.534***	0.434***
	(0.119)	(0.115)	(0.079)	(0.083)
Age	0.051***	0.050***	0.032**	0.030**
	(0.013)	(0.012)	(0.015)	(0.015)
Age sq	-0.000***	-0.000***	-0.000*	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Non-agricultural *hukou* at age 12	0.080**	0.087**	-0.001	0.012
	(0.035)	(0.035)	(0.039)	(0.039)
Mother Primary Edu	0.048	0.051	0.033	0.030
	(0.039)	(0.038)	(0.049)	(0.048)
Mother Junior High Edu	0.069	0.080	0.147***	0.136***
	(0.050)	(0.050)	(0.052)	(0.051)
Mother Senior High Edu	-0.040	-0.028	0.144**	0.141**
	(0.055)	(0.054)	(0.062)	(0.062)
Mother College+ Edu	0.221*	0.221*	0.214**	0.212**
	(0.118)	(0.116)	(0.100)	(0.099)
Urban	0.146***	0.161***	0.227***	0.258***
	(0.045)	(0.045)	(0.067)	(0.067)
East	0.367***	0.362***	0.436***	0.426***
	(0.030)	(0.030)	(0.035)	(0.035)
Mother's Year of Birth	0.009***	0.009***	-0.002	-0.003
	(0.003)	(0.003)	(0.004)	(0.004)
Managerial/Professional Occupations		0.171***		0.166***
		(0.034)		(0.038)
Employer		0.083		-0.001
		(0.107)		(0.174)

Constant	-11.554**	-10.854*	10.646	11.150
	(5.759)	(5.683)	(7.133)	(7.025)
Observations (person-waves)	3402	3402	2395	2395

Note: Robust standard errors in parentheses. Clustering at the individual level for pooled OLS. * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$. OrdinUG and KeyUG stand for Ordinary and Key Universities, respectively. Other controls include dummies for survey years.

Table A2: Full Set of IPWRA Treatment Effects of Table 4, Men

Second-step: log earnings equation	No HE	College STEM	College LEM	College Other	OrdinUG STEM	OrdinUG LEM	OrdinUG Other	KeyUG STEM	KeyUG LEM	Key Other
Age	0.033**	0.086	0.002	-0.020	0.062	-0.010	0.118**	0.246***	0.045	-0.118*
	(0.014)	(0.057)	(0.037)	(0.059)	(0.081)	(0.076)	(0.055)	(0.055)	(0.076)	(0.071)
Age sq	-0.000**	-0.001	0.000	0.000	-0.000	0.000	-0.001**	-0.003***	0.000	0.001*
	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Non-agr *hukou* at 12	0.032	0.122	0.095	0.148	0.213	0.114	-0.078	0.455***	0.125	0.340
	(0.043)	(0.113)	(0.068)	(0.136)	(0.273)	(0.174)	(0.122)	(0.148)	(0.155)	(0.207)
Mother Primary Edu	0.079*	-0.004	-0.044	-0.032	-0.153	0.069	-0.200	0.156	1.047***	0.096
	(0.042)	(0.120)	(0.094)	(0.104)	(0.216)	(0.176)	(0.156)	(0.138)	(0.189)	(0.186)
Mother Junior High	0.154***	0.144	-0.009	-0.229*	-0.541	-0.145	0.060	0.001	1.281***	-0.026
Edu	(0.051)	(0.122)	(0.110)	(0.137)	(0.583)	(0.378)	(0.153)	(0.163)	(0.249)	(0.294)
Mother Senior High	-0.034	0.116	-0.163	0.154	-0.217	-0.175	0.222	-0.248	1.301***	1.168***
Edu	(0.083)	(0.177)	(0.107)	(0.270)	(0.395)	(0.211)	(0.317)	(0.186)	(0.193)	(0.313)
Mother College/Uni	0.075	-0.015	0.238**	1.410	0.190	0.263*	0.783	-0.215	1.331***	0.351*
	(0.168)	(0.204)	(0.103)	(0.968)	(0.914)	(0.141)	(0.769)	(0.237)	(0.398)	(0.197)
Urban	0.071	0.166	0.350***	0.212**	0.213	-0.008	0.045	-0.359**	-2.211***	0.279
	(0.046)	(0.127)	(0.118)	(0.104)	(0.327)	(0.246)	(0.112)	(0.162)	(0.459)	(0.270)
East	0.305***	0.519***	0.480***	-0.021	0.632***	0.499***	0.499***	0.402***	0.564***	0.077
	(0.036)	(0.086)	(0.071)	(0.112)	(0.189)	(0.145)	(0.160)	(0.145)	(0.196)	(0.194)
Constant	6.488***	5.465***	7.047***	7.621***	5.709***	7.509***	5.093***	2.928***	7.084***	9.650***
	(0.275)	(1.002)	(0.781)	(1.172)	(1.377)	(1.439)	(1.177)	(1.035)	(1.390)	(1.483)

Table A2 (Continued)

First-step: Treatment equation	College STEM	College LEM	College Other	OrdinUG STEM	OrdinUG LEM	OrdinUG Other	KeyUG STEM	KeyUG LEM	Key Other
School cohort	0.026*	0.021*	-0.041**	0.027	0.018	0.013	0.072***	0.009	0.027
	(0.014)	(0.012)	(0.019)	(0.018)	(0.021)	(0.020)	(0.019)	(0.033)	(0.034)
Mother's birth year	0.018	-0.016*	0.026	-0.002	-0.006	0.007	-0.012	0.027	-0.033
	(0.012)	(0.010)	(0.016)	(0.017)	(0.017)	(0.021)	(0.015)	(0.032)	(0.022)
Non-agricul *hukou* at 12	-0.064	0.111	-0.575***	-0.660***	0.041	-0.733***	0.096	-0.292	0.195
	(0.136)	(0.131)	(0.214)	(0.197)	(0.200)	(0.233)	(0.259)	(0.285)	(0.345)
Mother Primary Edu	0.455***	0.403***	-0.190	1.041***	0.285	0.779***	0.672***	1.056***	0.853**
	(0.163)	(0.144)	(0.259)	(0.242)	(0.237)	(0.286)	(0.254)	(0.401)	(0.379)
Mother Junior High Edu	0.515***	0.311*	0.654***	0.924***	0.497*	0.208	-0.616	0.552	0.157
	(0.180)	(0.186)	(0.248)	(0.277)	(0.292)	(0.378)	(0.411)	(0.476)	(0.495)
Mother Senior High Edu	0.984***	0.940***	0.860**	1.814***	1.015***	0.811*	0.333	1.949***	0.247
	(0.214)	(0.203)	(0.337)	(0.309)	(0.292)	(0.434)	(0.402)	(0.417)	(0.644)
Mother College/Uni	2.775***	2.255***	2.113***	1.785**	0.268	2.616***	2.800***	3.382***	1.641
	(0.407)	(0.410)	(0.581)	(0.796)	(1.079)	(0.641)	(0.538)	(0.647)	(1.121)
Urban	0.527***	0.823***	0.145	0.778***	2.661***	0.650**	0.746**	2.144***	0.021
	(0.170)	(0.181)	(0.210)	(0.259)	(0.585)	(0.303)	(0.316)	(0.731)	(0.415)
East	-0.570***	-0.259**	-0.292*	-0.203	-0.544***	-0.996***	0.390*	-0.658**	0.113
	(0.126)	(0.113)	(0.169)	(0.176)	(0.185)	(0.240)	(0.217)	(0.275)	(0.331)
Constant	-87.795***	-11.329	28.069	-53.152***	-30.592	-42.240**	-122.870***	-76.107***	5.965
	(14.475)	(12.363)	(18.735)	(17.373)	(20.250)	(21.362)	(22.453)	(26.431)	(33.143)

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. OrdinUG and KeyUG stand for Ordinary and Key Universities, respectively. Other controls include dummies for survey years. No HE is the omitted category in the treatment equation.

Table A3: Full Set of IPWRA Treatment Effects of Table 4, Women

Second-step: log earning equation	No HE	College STEM	College LEM	College Other	OrdinUG STEM	OrdinUG LEM	OrdinUG Other	KeyUG STEM	KeyUG LEM	Key Other
Age	0.021	0.028	0.009	0.060	0.057	-0.077	0.147**	0.115	-0.078	-0.165**
	(0.018)	(0.047)	(0.029)	(0.053)	(0.084)	(0.068)	(0.071)	(0.155)	(0.072)	(0.065)
Age sq	-0.000	0.000	0.000	-0.000	-0.000	0.001	-0.002*	-0.001	0.002*	0.002***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Non-agr *hukou* at 12	-0.013	-0.177*	-0.026	0.099	0.279**	0.023	-0.031	0.546**	0.495***	-0.177*
	(0.050)	(0.101)	(0.074)	(0.115)	(0.114)	(0.269)	(0.131)	(0.245)	(0.107)	(0.106)
Mother Primary Edu	0.077	0.317***	0.045	-0.245*	0.039	-0.547	-0.465	-0.816***	0.006	0.366***
	(0.054)	(0.118)	(0.090)	(0.148)	(0.177)	(0.587)	(0.348)	(0.175)	(0.222)	(0.121)
Mother Junior High Edu	0.104*	0.364**	-0.052	0.311**	0.115	-0.293	0.130	-0.007	0.030	0.393***
	(0.061)	(0.143)	(0.105)	(0.151)	(0.173)	(0.400)	(0.178)	(0.125)	(0.245)	(0.145)
Mother Senior High Edu	0.123	0.440***	0.011	0.104	0.049	0.087	-0.073	-0.786**	-0.154	0.316
	(0.081)	(0.162)	(0.110)	(0.225)	(0.204)	(0.333)	(0.205)	(0.394)	(0.281)	(0.215)
Mother College/Uni	0.155	0.578*	0.077	-0.002	0.207	0.395	-0.055	-0.591	-0.390*	0.623***
	(0.149)	(0.297)	(0.196)	(0.296)	(0.210)	(0.446)	(0.234)	(0.543)	(0.226)	(0.208)
Urban	0.403***	0.088	0.230**	0.028	-0.448***	0.166	0.692	0.311*	0.153	0.552***
	(0.082)	(0.152)	(0.109)	(0.143)	(0.169)	(0.346)	(0.649)	(0.180)	(0.215)	(0.122)
East	0.332***	0.590***	0.659***	0.329***	0.621***	0.595***	0.267*	0.776***	0.157	0.314***
	(0.043)	(0.096)	(0.064)	(0.115)	(0.120)	(0.135)	(0.142)	(0.108)	(0.108)	(0.096)
Constant	6.203***	5.750***	6.429***	5.727***	5.865***	9.203***	4.282***	4.509	7.831***	9.653***
	(0.320)	(0.829)	(0.515)	(0.917)	(1.442)	(1.680)	(1.498)	(2.759)	(1.341)	(1.340)

Table A3 (Continued)

First-step: Treatment equation	College STEM	College LEM	College Other	OrdinUG STEM	OrdinUG LEM	OrdinUG Other	KeyUG STEM	KeyUG LEM	Key Other
School cohort	-0.003	0.019	0.020	0.106***	0.033	0.074***	0.028	0.097***	0.008
	(0.018)	(0.016)	(0.021)	(0.024)	(0.025)	(0.023)	(0.039)	(0.033)	(0.031)
Mother's birth year	0.033**	-0.006	-0.003	-0.072***	-0.001	-0.019	0.028	-0.031	0.008
	(0.015)	(0.012)	(0.017)	(0.018)	(0.021)	(0.020)	(0.029)	(0.031)	(0.024)
Non-agricul *hukou* at 12	-0.139	0.158	-0.315*	-0.379	0.606**	-0.317	-0.425	0.994***	-0.765***
	(0.172)	(0.135)	(0.187)	(0.263)	(0.252)	(0.218)	(0.320)	(0.356)	(0.258)
Mother Primary Edu	0.349	0.615***	0.746***	0.588*	0.081	0.630**	0.856	0.117	0.395
	(0.216)	(0.175)	(0.232)	(0.353)	(0.305)	(0.304)	(0.694)	(0.448)	(0.403)
Mother Junior High Edu	0.835***	0.922***	1.091***	1.622***	0.796***	0.569	1.802***	0.315	1.547***
	(0.239)	(0.193)	(0.252)	(0.391)	(0.290)	(0.355)	(0.681)	(0.435)	(0.365)
Mother Senior High Edu	1.144***	1.364***	0.786**	1.828***	0.978***	1.634***	2.568***	1.394***	2.148***
	(0.270)	(0.224)	(0.354)	(0.469)	(0.349)	(0.354)	(0.649)	(0.424)	(0.374)
Mother College/Uni	1.089**	1.536***	0.679	2.465***	-0.258	2.687***	4.050***	2.320***	2.355***
	(0.540)	(0.414)	(0.765)	(0.661)	(1.058)	(0.488)	(0.759)	(0.592)	(0.719)
Urban	0.328	0.366*	-0.092	0.413	0.793	0.910**	1.398*	-0.400	0.823*
	(0.246)	(0.207)	(0.234)	(0.347)	(0.488)	(0.359)	(0.741)	(0.455)	(0.434)
East	-0.471***	0.243**	-0.384**	-0.213	-0.351*	-0.957***	0.395	0.996***	-0.001
	(0.159)	(0.120)	(0.179)	(0.229)	(0.213)	(0.215)	(0.323)	(0.284)	(0.238)
Constant	-61.119***	-28.941*	-34.366	-73.006**	-66.468***	-111.785***	-115.734***	-135.868***	-36.276
	(20.464)	(15.577)	(23.651)	(29.561)	(21.969)	(22.792)	(38.185)	(28.627)	(26.200)

Note: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. OrdinUG and KeyUG stand for Ordinary and Key Universities, respectively. Other controls include dummies for survey years. No HE is the omitted category in the treatment equation.

Table A4: Higher Education Subjects and Tiers Choice, by family background

	By HE Subjects				By HE Tiers			
%	**Senior High School**	**STEM**	**LEM**	**Other**	**Senior High School**	**College**	**OrdinUG**	**KeyUG**
By mother's education (share in parentheses), Agricultural *hukou* at 12, N=1,489 (52.9%)								
Illiterate (45.8%)	65.1	11.3	13.9	9.7	65.1	23.6	7.5	3.8
Primary (30.8%)	49.2	17.7	18.5	14.6	49.2	30.7	12.6	7.4
Junior High (17.7%)	48.5	23.9	14.8	12.9	48.5	30.7	12.9	8.0
Senior High (5.2%)	40.3	22.1	24.7	13.0	40.3	33.8	20.8	5.2
College or above (0.5%)	0	42.9	14.3	42.9	0	57.1	14.3	28.6
By mother's education (share in parentheses), Non-agricultural *hukou* at 12, N=1,324 (47.1%)								
Illiterate (19.9%)	66.9	9.9	17.1	6.1	66.9	20.9	8.4	3.8
Primary (23.1%)	51.6	17.3	22.9	8.2	51.6	30.1	10.5	7.8
Junior High (28.0%)	39.9	19.7	28.0	12.4	39.9	37.4	16.2	6.5
Senior High (23.3%)	27.2	24.3	34.3	14.2	27.2	38.2	20.7	13.9
College or above (5.7%)	17.3	29.3	36.0	17.3	17.3	38.7	21.3	22.7

Note: Distinct individuals in Wave 1. OrdinUG and KeyUG stand for Ordinary and Key universities, respectively.